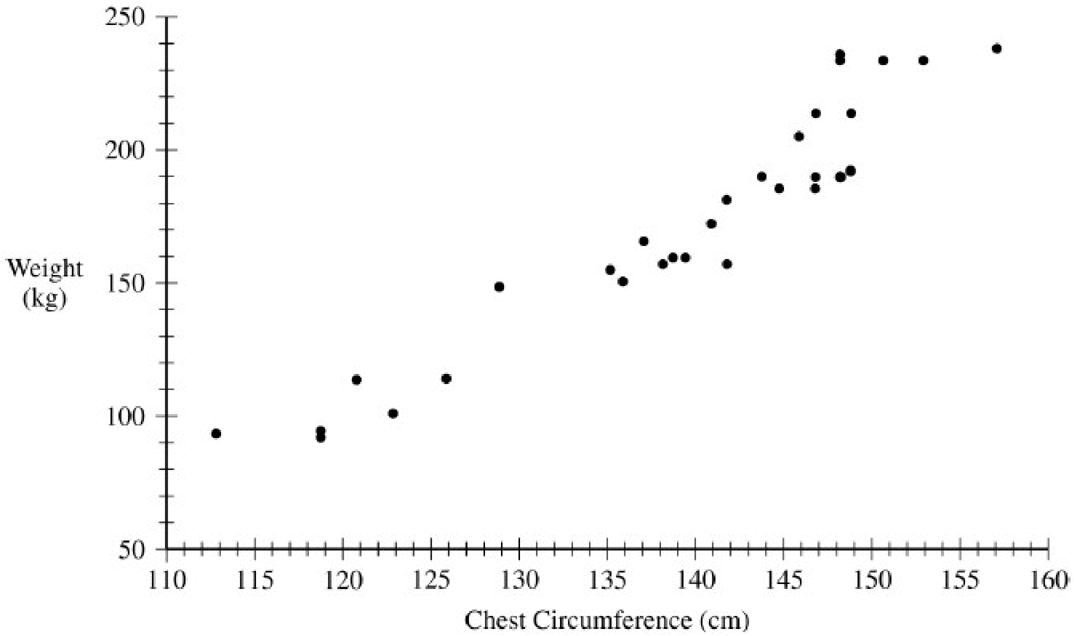
• Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.



(A) Describe the relationship between chest circumference and weight of male tule elk in context.

Following is the equation of the least-squares regression line relating chest circumference and weight for male tule elk.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

(B) The weight of one male tule elk with a chest circumference of 145.9 cm is 204.3 kg.

(i) Using the equation of the least-squares regression line, calculate the predicted weight for this male tule elk. Show your work.

(ii) Calculate the residual for this male tule elk. Show your work.

The equation of the least-squares regression line relating chest circumference and weight for male tule elk is repeated here.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

(C) Interpret the slope of the least-squares regression line in context.

(D) The sambar, another species of deer, is similar in size to the tule elk. The slope of the population regression line relating chest circumference and weight for all male sambars is 4.5 kilograms per centimeter. A wildlife biologist wants to determine whether the slope of the population regression line for male tule elk is different than that for male sambars. Let β represent the slope of the population regression line for male tule elk. The wildlife biologist conducted a test of the following hypotheses using the sample of 30 tule elk.

$$\mathrm{H_0} : \beta = 4.5$$
$$\mathrm{H_a} : \beta \neq 4.5$$

The test statistic was calculated to be 3.408. Assume all conditions for inference were met.

(i) Determine the p-value of the test.

(ii) At a significance level of $\alpha = 0.05$, what conclusion should the wildlife biologist make regarding the slope of the population regression line for male tule elk? Justify your response.

Solution:

(A) The scatterplot reveals a strong, positive, roughly linear association between the chest circumference and weight of tule elk. There are no points that seriously deviate from the straight-line pattern of the points in the plot.

(B) (i) The predicted weight of a male tule elk with a chest circumference of 145.9 cm is $-350.3 + 3.7455(145.9) \approx 196.17$ kg.
(ii) The residual for a male tule elk with a chest circumference of 145.9 cm with an actual weight of 204.3 kg is $204.3 - 196.17 \approx 8.13$ kg.

(C) The value of the slope of the least-squares regression line is 3.7455. This value indicates that the predicted weight of a tule elk increases by 3.7455 kilograms for each additional centimeter of chest circumference.

(D) (i) The degrees of freedom for the test of slope are $n - 2 = 30 - 2 = 28$. The t-table shows that for 28 degrees of freedom, the p-value for a one-sided test would be 0.001. Because this is a two-sided test, the p-value is $(2)(0.001) = 0.002$.
(ii) Because the p-value $= 0.002$ is less than $\alpha = 0.05$, reject the null hypothesis. There is sufficient statistical evidence that the population slope for the linear regression of weight vs. chest circumference for male tule elk is different from 4.5 kg/cm.

• Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000
$S = 0.237$	$R\text{-}Sq = 0.873$		$R\text{-}Sq$ (adj) $= 0.868$	

(A) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

(B) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

(C) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

(D) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

Solution

(A) The equation of the least squares regression line is

$$\widehat{\text{electricity production}} = 0.137 + 0.240 \times \text{wind velocity}.$$

(B) The slope coefficient of 0.240 indicates that for each additional mph of wind speed, the expected electricity production increases by 0.240 amperes. Thus, the expected electricity production is $10 \times 0.240 = 2.40$ amperes higher on a day with 25 mph wind velocity as compared to a day with 15 mph wind velocity.

(C) The proportion of variation in electricity production that is explained by the linear relationship with wind speed is R^2, which the regression output reports to be 0.873.

(D) Yes, there is very strong statistical evidence that the population slope differs from zero, so electricity production is linearly related to wind speed. For testing the hypotheses $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$, where β represents the population slope, the output reveals that the test statistic is $t = 12.63$ and the p-value (to three decimal places) is 0.000. Because the p-value is so small (much less than both 0.05 and 0.01), the sample data provide very strong statistical evidence that electricity production is linearly related to wind speed.

Problems adapted from the College Board released tests.